TITLE

UTILIZATION OF STARCH PRODUCTS FOR BIOLOGICAL PRODUCTION BY FERMENTATION

This application claims the benefit of U.S. Provisional Application No. 60/405896, filed August 23, 2002.

FIELD OF THE INVENTION

The present invention relates to the field of molecular biology. More specifically it describes microbial hosts containing genes that express enzymes that effectively convert starch products into a fermentation product.

BACKGROUND OF THE INVENTION

Fermentation is an important technology for the biocatalytic conversion of renewable feedstocks into desirable products. Carbohydrates are traditional feedstocks in the fermentation industry. It is often the case that carbohydrates used as a substrate contribute more to the cost of manufacture than any other single component. Depending on the particular process, from 25 to 70 % of the total cost of fermentation may be due to the carbohydrate source. (Crueger and Crueger, *Biotechnology: A Textbook of Industrial Microbiology*, Sinauer Associates: Sunderland, MA., pp 124-174 (1990); Atkinson and Mavituna, *Biochemical Engineering and Biotechnology Handbook*, 2nd ed.; Stockton Press: New York, pp 243-364 (1991)). For such economic reasons, highly purified glucose or sucrose can seldom be used as a substrate.

Starch, a carbohydrate, is a mixture of two different polysaccharides each consisting of chains of linked, repeating monosaccharide (glucose) units. The mixture consists of two separate polysaccharides, amylose and amylopectin. Amylose is a linear polysaccharide with glucose units connected exclusively through α(1,4) glycosidic linkages. Glucose units in amylopectin are also linked through α(1,4) glycosidic linkages, and additionally are linked through α(1,6) glycosidic linkages, about one every 30 glucose residues. The ratio of amylopectin to amylose in starch varies from one plant species to another, but is generally in the range of 3-4 to 1 (Kainuma, pp 125-150 in Starch; Whistler, Bemiller, and Pashcall eds., Academic Press, Orlando, FL (1984)).

Commercial starch is produced primarily through the wet milling process. The final products from a wet mill, however, include very little unprocessed starch. By far, the majority of products made are in the form of fully processed starch (monosaccharides, including glucose) or smaller

degradation products derived from starch. Typically, an amylase enzyme is used to break starch into smaller chains (Blanchard, Technology of Corn Wet Milling (1992), Elseiver, Amsterdam, The Netherlands, pp. 174-215). Various commercial sources of α-amylase exist, but, regardless of enzyme source, reaction products are generally the same with respect to size and linkage-type. Amylase digestion of starch results in a product known as a limit dextrin that includes small starch chains containing 2 – 10 glucose units (oligosaccharides). Because amylase cannot hydrolyze the α(1,6) glycosidic linkages in amylopectin, limit dextrins contain both α(1,4)- and α(1,6)-linked glucose oligosaccharides. Alternatively, raw starch may be treated by non-enzymatic means (for example, by acid hydrolysis) to produce starch products substantially similar to limit dextrin.

In the wet milling industry, limit dextrins are further processed into glucose for use as a carbon source for fermentations to produce various chemicals, commercial enzymes, or antibiotics. Relatively pure glucose is preferred as a carbohydrate source when the popular biocatalyst, *Escherichia coli*, is used in the fermentation process. This is because *E. coli* does not utilize components of limit dextrins (i.e., panose, isomaltose, and high molecular weight oligosaccharides with chains larger than about ten α(1,4)-linked glucose units) that are commonly contained in alternate low-cost fermentation media (Lin, *Escherichia coli* and *Salmonella typhimuium*, pp. 245-265, Neidhardt, ed.; American Society for Microbiology, Washington, D. C. (1987)). Glucose oligomers containing α(1,6)-linkages are not transported into the cell and *E. coli* does not produce an enzyme that degrades this material when supplied extracellularly (Palmer *et al.*, *Eur. J. Biochem.* 39:601-612 (1973)).

Making relatively pure glucose from starch that is suitable for use by *E. coli* requires many process steps and additional enzymes, adding significantly to the cost of product manufacture.

Thus, the problem to be solved is the lack of a process to utilize low-cost starch products in large-scale fermentative production processes. An ability to more completely ferment low cost, partially degraded starch would lower the cost of manufacture for products made through fermentation.

SUMMARY OF THE INVENTION

Applicants have provided an isolated nucleic acid molecule encoding an α(1,6)-linked glucose oligosaccharide hydrolyzing enzyme selected from the group consisting of: (a) an isolated nucleic acid molecule

encoding the amino acid sequence selected from the group consisting of SEQ ID NOs:2, 4, and 6; (b) a nucleic acid molecule that hybridizes with (a) under the following hybridization conditions: 0.1X SSC, 0.1 % SES, 65°C and washed with 2X SSC, 0.1 % SDS followed by 0.1X SSC, 0.1 % SDS; and (c) a nucleic acid molecule that is complementary to (a) or (b).

Applicants have provided nucleic acid compositions comprising coding regions for a signal peptide and an α(1,6)-linked glucose oligosaccharide hydrolyzing enzyme such that a chimeric protein is expressed that directs the hydrolyzing activity external to the cytoplasm (extracellularly). The isolated nucleic acid molecule may encode a signal peptide as set forth in SEQ ID NO:24 or SEQ ID NO:25. The nucleic acid sequence of the signal sequence is SEQ ID NO:26 or SEQ ID NO:27. The isolated nucleic acid molecule may encode an α(1,6)-linked glucose oligosaccharide hydrolyzing polypeptide as set forth in SEQ ID NOs:2, 4, 6, 17, or 31.

Applicants have provided recombinant organisms comprising an α(1,6)-linked glucose oligosaccharide hydrolyzing enzyme that enables the utilization of exogenously added α(1,6)-linked glucose oligosaccharides (e.g., isomaltose and panose) for the fermentative production of useful products. The α(1,6)-linked glucose oligosaccharide hydrolyzing polypeptide may be selected from SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:17, or SEQ ID NO:31. The invention also encompasses an α(1,6)-linked glucose oligosaccharide hydrolyzing polypeptide encoded by the nucleic acid molecule set forth in SEQ ID NOs:1, 3, 5, 16, or 30. The invention also includes isolated nucleic acid molecules selected from the group consisting of SEQ ID NO:3, SEQ ID NO:28, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, or SEQ ID NO:42. The invention also includes the polypeptide SEQ ID NO:4, SEQ ID NO:29, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, and SEQ ID NO:43.

The invention also encompasses a chimeric gene comprising the isolated nucleic acid molecules set forth herein operably linked to suitable regulatory sequences. The suitable regulatory sequence is selected from the group comprising *CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI, AOX1, lac, ara, tet, trp,* IP_L, IP_R, *T7, tac, trc, apr, npr, nos*, and GI. The invention encompasses transformed host cells wherein the chimeric gene is integrated into the chromosome or is plasmid-borne.

Applicants have also provided a method for degrading limit dextrin comprising:

(a) contacting a transformed host cell comprising:

(i) a nucleic acid molecule encoding the enzymes selected from the group consisting of SEQ ID NOs:2, 6, 17 and 31;

(ii) a nucleic acid molecule that hybridizes with (i) under the following hybridization conditions: 0.1X SSC, 0.1 % SDS, 65°C and washed with 2X SSC, 0.1 % SDS followed by 0.1X SSC, 0.1 % SDS; or

(iii) a nucleic acid molecule that is complementary to (i) or (ii),

with an effective amount of limit dextrin substrate under suitable growth conditions; and

(b) optionally recovering the product of step (a).

The invention also encompasses a method for producing a target molecule in a recombinant host cell comprising: contacting a transformed host cell comprising: (i) an isolated nucleic acid molecule encoding a chimeric protein comprised of a signal peptide linked to an α(1,6)-linked glucose oligosaccharide hydrolyzing polypeptide; (ii) a nucleic acid molecule that hybridizes with (i) under the following hybridization conditions: 0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS; or (iii) a nucleic acid molecule that is complementary to (i) or (ii); and a chimeric gene for converting mononsaccharides to the target molecule, in the presence of limit dextrin under suitable conditions whereby the target molecule is produced; and optionally recovering the target molecule produced. The signal peptide may be selected from SEQ ID NO:24 or SEQ ID NO:25. The α(1,6)-linked glucose oligosaccharide hydrolyzing polypeptide may be selected from SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:17 or SEQ ID NO:31. The transformed host cell may be selected from bacteria, yeast or filamentous fungi. This invention includes producing 1,3 propanediol, glycerol, and cell mass from limit dextrin.

The invention also encompasses a polypeptide having an amino acid sequence that has at least 69% identity based on the BLASTP method of alignment when compared to a polypeptide having the sequence as set forth in SEQ ID NO:17, the polypeptide having an α(1,6)-linked glucose oligosaccharide hydrolyzing activity.

BRIEF DESCRIPTION OF THE DRAWINGS, BIOLOGICAL DEPOSITS, AND SEQUENCE DESCRIPTIONS

Figures 1a through 1d show the results of the *E. coli* strain DH5a containing the plasmids pUC18 (Fig. 1a) (negative control) and pUC18 containing the mature coding sequence from the clones j20 (Fig. 1b), k1 (Fig. 1c), or h12 (Fig.1d). Total protein extracts were isolated from sonicated cells and incubated with panose (250 μg/ml) at 37 °C for two hours. A high performance anion exchange chromatogram of the products after digestion is shown.

Applicants made the following biological deposits under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purposes of Patent Procedure at the American Type Culture Collection (ATCC) 10801 University Boulevard, Manassas, VA 20110-2209:

Depositor Identification Reference	Int'l. Depository Designation	Date of Deposit
Escherichia coli RJ8n	ATCC PTA-4216	9 April 2002

The listed deposit(s) will be maintained in the indicated international depository for at least thirty (30) years and will be made available to the public upon the grant of a patent disclosing it. The availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by government action.

Applicants provide a sequence listing containing 43 sequences. The sequences are in conformity with 37 C.F.R. 1.821 - 1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions) and with the corresponding United Stats Patent and Trademark Office Rules set forth in 37 C.F.R. §1.822.

ORF Name	Gene Name	SEQ ID Base	SEQ ID Peptide	Strain of Origin
mbc1g.pk007.h12	*algB*	1	2	*Bifidobacterium breve*
mbc2g.pk018.j20	*algA*	3	4	*Bifidobacterium breve*
mbc1g.pk026.k1	*algA*	5	6	*Bifidobacterium breve*
dexB	*dexB*	16	17	*Streptococcus mutans*

SEQ ID NOs:1-6 are nucleic and amino acid sequences of three genes/gene products obtained from *Bifidobacterium breve* ATCC 15700.

SEQ ID NOs:7-15 and 18-23 are primers for PCR.

SEQ ID NOs:16-17 are nucleic and amino acid sequences disclosed in public databases for *Streptococcus mutans* (ATCC 25175D).

SEQ ID NO:24 is the amino acid sequence for the native signal peptide from the *Bifidobacterium breve* gene, mbc2g.pk018.j20 (also contained within SEQ ID NO:3).

SEQ ID NO:25 is the amino acid sequence for the non-native signal peptide used to target enzymes coded for by the *Bifidobacterium breve* mbc1g.pk026.k1 and *Streptococcus mutans* dexB genes.

SEQ ID NO:26 is the nucleic acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 signal peptide

SEQ ID NO:27 is the nucleic acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide.

SEQ ID NO:28 is the nucleic acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 signal peptide linked to the coding sequence for the *Bifidobacterium breve* mbc2g.pk018.h12 gene.

SEQ ID NO:29 is the amino acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 signal peptide linked to the amino acid sequence for the *Bifidobacterium breve* mbc2g.pk018.h12 gene.

SEQ ID NO:30 is the nucleic acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 without its native signal peptide sequence.

SEQ ID NO:31 is the amino acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 without its native signal peptide sequence.

SEQ ID NO:32 is the nucleic acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 signal peptide linked to the coding sequence for the *Bifidobacterium breve* mbc2g.pk018.k1 gene.

SEQ ID NO:33 is the amino acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 signal peptide linked to the amino acid sequence for the *Bifidobacterium breve* mbc2g.pk018.k1 gene.

SEQ ID NO:34 is the nucleic acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 signal peptide linked to the coding sequence for the *Streptococcus mutans* dexB gene.

SEQ ID NO:35 is the amino acid sequence for the *Bifidobacterium breve* gene mbc2g.pk018.j20 signal peptide linked to the amino acid sequence for the *Streptococcus mutans* dexB gene.

SEQ ID NO:36 is the nucleic acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide linked to the coding sequence for the *Bifidobacterium breve* mbc2g.pk018.h12 gene.

SEQ ID NO:37 is the amino acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide linked to the amino acid sequence for the *Bifidobacterium breve* mbc2g.pk018.h12 gene.

SEQ ID NO:38 is the nucleic acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide linked to the coding sequence for the *Bifidobacterium breve* mbc2g.pk018.j20 gene.

SEQ ID NO:39 is the amino acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide linked to the amino acid sequence for the *Bifidobacterium breve* mbc2g.pk018.j20 gene.

SEQ ID NO:40 is the nucleic acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide linked to the coding sequence for the *Bifidobacterium breve* mbc2g.pk018.k1 gene.

SEQ ID NO:41 is the amino acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide linked to amino acid sequence for the *Bifidobacterium breve* mbc2g.pk018.k1 gene.

SEQ ID NO:42 is the nucleic acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide linked to the coding sequence for the *Streptococcus mutans* dexB gene.

SEQ ID NO:43 is the amino acid sequence for the *Bacillus subtilis* neutral protease gene signal peptide linked to amino acid sequence for the *Streptococcus mutans* dexB gene.

DETAILED DESCRIPTION OF THE INVENTION

Applicants have solved the stated problem. The present invention provides several enzymes that, when expressed in a production host, enable the host to utilize α(1,6)-linked glucose oligosaccharides, which are components of low cost starch products. The invention also provides signal sequences that enable α(1,6)-linked glucose oligosaccharide hydrolyzing enzymes to be targeted extracellularly.

Low cost starch products are obtained, for example, from the action of commercially available amylase enzymes on raw starch and other feed stocks containing α(1,6)-linked glucose oligosaccharides to produce a limit dextrin. The efficient use of the low cost starch products requires genetically engineering a host organism (for example, *E. coli*), such that the recombinant organism produces an enzyme that degrades α(1,6)-linked glucose oligosaccharides. Enzymes that degrade α(1,6)-linked

glucose oligosaccharides are known (Vihinen and Mantsala *Crit. Rev. in Biochem. Mol. Biol.* 4:329-427 (1989)). Further, enzymes that degrade these linkages are known to be present both intracellularly (within the cytoplasm) and extracellularly (external to the cytoplasm) in their native state.

Where a host organism lacks a transport system, engineering an intracellular enzyme to have access to limit dextrin (or other feedstocks containing α(1,6)-linked glucose oligosaccharides) supplied externally may be accomplished by adding a native or non-native signal peptide. Signal peptides enable the α(1,6)-linked glucose oligosaccharide degrading protein to be directed to an extracellular location (external to the cytoplasm), and give access to substrates not taken into the cell (Nagarajan *et al., Gene* 114:121-126 (1992)). Examples of signal peptides that translocate protein across the cell's membrane include, but are not limited to, SEQ ID NOs:24 and 25. Proteins containing a signal peptide are directed to the secretory pathway and are then translocated across the cell's membrane. The general mechanism of protein secretion is conserved among all gram-negative and gram-positive bacteria (Simonen and Palva (1993) *Microbiol. Rev.* 57:109-137; Fekkes and Driessen (1999) *Microbiol. Rev.* 63:161-173). All bacterial signal peptides contain a string of 13 to 20 hydrophobic amino acids (Bae and Schneewind, *J. Bacteriol.,* 185:2910-2919 (2003)).

Native *E. coli* does not hydrolyze α(1,6)-glycosidic linkages, thus the compounds containing (1,6)-linkages are not utilized in fermentations. The (1,6)-linkages are hydrolyzed by both "isoamylase" and "glucosidase" enzymes (isomaltose and panose are model compounds for (1,6)-linked oligosaccharides). A recombinant *E. coli* containing a non-native extracellular "isoamylase" or "glucosidase" will utilize compounds containing (1,6)-linkages (e.g., isomaltose and panose) in fermentations to produce useful products. Further, any recombinant organism containing a non-native extracellular "isoamylase" or "glucosidase" will utilize compounds containing (1,6)-linkages more efficiently. Increased utilization efficiency will be through constitutive expression or altered timing of the recombinant "isoamylase" or "glucosidase" genes. Recombinant gene expression will also increase the level of activity over that of any endogenous "isoamylase" or "glucosidase" genes that may be present, thus increasing utilization of (1,6)-linked substrate.

The present invention may be used to produce various products of biofermentation including, but not limited to, organic acids, antibiotics, amino acids, enzymes, vitamins, alcohols such as bioethanol, and cell mass. The bio-production of glycerol, 1,3-propanediol, and cell mass using limit dextrin made available as a carbon source to the host microorganism through use of the signal peptide serve to exemplify the invention.

The polyol, 1,3-propanediol, is a monomer useful for producing polyester fibers and manufacturing polyurethanes and cyclic compounds. A process for the biological production of 1,3-propanediol by a single organism from carbon substrate such as glucose or other sugars has been described in U. S. Patent No. 5,686,276, incorporated by reference herein.

Starch is a homopolysaccharide of glucose. It is synthesized in higher plants as a granule containing two components, amylose and amylopectin (Vihinen and Mantsala, *Crit. Rev. Biochem. Mol. Biol.*, 24:329-418 (1989)). Amylose, essentially a linear polysaccharide formed by α(1,4)-linked glucose residues, accounts for 15-25 % of the granule (content varies with plant species). By contrast, amylopectin is highly branched, with about 4 to 5 % of the glucosidic linkages being α(1,6)-linked glucose residues. Amylolytic enzymes that degrade starch are well studied. Metabolism of starch, by first degrading the polymer into individual glucose residues in higher plant species, requires the interaction of several amylolytic enzymes.

Amylolytic enzymes, acting alone, often only partially degrade starch into smaller linear or branched chains. Combinations of amylolytic enzymes or enzyme combinations along with acid treatment may be used to increase the depolymerization of starch.

Enzymes and enzyme combinations may degrade starch partially, resulting in smaller linear or branched chains, or completely to glucose. The α-glucosidases hydrolyze both (1,4)- and (1,6)-linkages found in oligosaccharides which are formed by the action of other amylolytic enzymes such as α-amylases, β-amylases, glucoamylases, isoamylases and pullulanases, or by acid and heat treatments.

α-Glucosidases (α-D-glucoside glucohydrolase; for example, EC 3.2.1.20) are distributed widely among microorganisms. They hydrolyze (1,4)- and (1,6)-linkages and liberate α-D-glucose units from the nonreducing end. Various types of these enzymes with different (and wide) substrate specificity have been found in bacterial species of the genus

Bacillus, *Streptococcus*, *Escherichia*, *Pseudomonas*, hyperthermophilic archaeobacteria such as *Pyrococcus*, *Thermococcus,* and *Thermotoga*, and fungal species such as *Penicillium*, *Tetrahymena*, *Saccharomyces,* and *Aspergillus*.

The enzyme from *Aspergillus niger* has been intensively studied for many years and possesses wide substrate specificity. It hydrolyzes such substrates such as maltose, kojibiose, nigerose, isomaltose, phenyl-α-glucoside, phenyl-α-maltoside, oligosaccharides, maltodextrin, and soluble starch. Similar properties are exhibited by α-glucosidases from *A. oryzae*, *Bacillus subtilis*, and *B. cereus* and the hyperthermophilic archaea.

Oligo-(1,6)-glucosidase or isomaltase (dextrin 6-α-D-glucanohydrolase, EC 3.2.1.10; coded for by the dexB gene) is an enzyme similar to α-glucosidase (Krasikov *et al.*, *Biochemistry* (Moscow). 66:332-348 (2001)). It catalyzes the hydrolysis of (1,6)-α-D-glucosidic linkages in isomaltose and dextrins produced from starch and glycogen by α-amylase (Enzyme Nomenclature, C. Webb, ed. (1984) Academic Press, San Diego, CA.). The enzyme is less well distributed than the α-glucosidases, but is found in organisms such as *Bacillus* species including *B. thermoglucosidius* KP1006, *B. cereus* ATCC 7064, and possibly *B. amyloliquefaciens* ATCC 23844 (Vihinen and Mantsala, Critical Reviews in Biochemistry. 24:329-418 (1989)), as well as *Bacillus coagulans* (Suzuki and Tomura, *Eur. J. Biochem.*, 158:77-83 (1986)). The *Bacillus* enzymes are typically 60-63 kDa in size. An oligo-(1,6)-alpha-glucosidase (EC 3.2.1.10) has also been isolated from *Thermoanaerobium* Tok6-B1, with a reported molecular mass of 30-33 kDa.

The dexB enzyme from *Steptococcus mutans* has a pattern of activity similar to the dextranase enzymes (EC 3.2.1.11) that catalyze the endohydrolysis of the (1,6)-α-D-glucosidic linkages in dextran. There is a high degree of similarity between the dexB enzyme and *Bacillus* spp. oligo-(1,6)-glucosidases (Whiting et al., *J. Gen. Microbiol.*, 139:2019-2026 (1993)). DexB is approximately 62 kDa in size (Aduse-Opoku et al., *J. Gen Microbiol.,* 137:757-764 (1991)).

Enzymes with α(1,6) hydrolase activity belong to a very broad category of over 81 recognized families of glucosyl hydrolases (Henrissat, *Biochem. J.*, 280:309-316 (1991); Henrissat and Bairoch, *Biochem. J.,* 293:781-788 (1993)). The broad grouping of enzymes capable of utilizing α(1,6) linked glucose units as a fermentable substrate is further emphasized by demonstrating the utility of this invention, using enzymes

with as little as 69 % amino acid sequence identity. Enzymes with the ability to depolymerize oligosaccharides containing α(1,6)-linked glucose residues are known and include glucoamylase, (EC 3.2.1.3, also known as amyloglucosidase), which rapidly hydrolyzes (1,6)-α-D-glucosidic bonds or linkages when the next linkage in sequence is a (1,4)-α-D-glucosidic linkage; α -dextrin endo-(1,6)-α-glucanosidase (EC 3.2.1.41, also known as pullulanase), which degrades (1,6)-α-D-glucosidic linkages in pullulan, amylopectin, glycogen, and the α- and β-amylase limit dextrins of amylopectin and glycogen; sucrase (EC 3.2.1.48), which is isolated from intestinal mucosa and has activity against isomaltose; isoamylase (EC 3.2.1.68), which hydrolyzes the (1,6)-α-D-glucosidic linkages in glycogen, amylopectin and their β-limit dextrins; and glucan (1,6)-α-glucosidase (EC 3.2.1.70), which hydrolyzes successive glucose residues from (1,6)-α-D-glucans and derived oligosaccharides.

In the context of this disclosure, a number of terms are used.

The term "starch" refers to a homopolysaccharide composed of D-glucose units linked by glycosidic linkages that forms the nutritional reservoir in plants. Starch occurs in two forms, amylose and amylopectin. In amylose, D-glucose units are linked exclusively by α(1,4) glycosidic linkages. Chains composed of multiple α(1,4) glycosidic linkages are considered to be linear or unbranched. In amylopectin, while the predominant connection is via α(1,4) glycosidic linkages, the occasional presence of an α(1,6) glycosidic linkage forms a branch point amongst the otherwise linear sections. Amylopectin contains about one α(1,6) linkage per thirty α(1,4) linkages.

The term "monosaccharide" refers to a compound of empirical formula $(CH_2O)_n$, where $n \geq 3$, the carbon skeleton is unbranched, each carbon atom except one contains a hydroxyl group, and the remaining carbon atom is an aldehyde or ketone at carbon atom 2. The term "monosaccharide" also refers to intracellular cyclic hemiacetal or hemiketal forms. The most familiar monosaccharide is D-glucose. The cyclic form of D-glucose involves reaction of the hydroxyl group of carbon atom 5 with the aldehyde of carbon atom 1 to form a hemiacetal, the carbonyl carbon being referred to as the anomeric carbon.

The terms "glycosidic bond" and "glycosidic linkage" refers to acetals formed by reaction of an anomeric carbon with a hydroxyl group of an alcohol. Reaction of the anomeric carbon of one D-glucose molecule with the hydroxyl group on carbon atom 4 of a second D-glucose molecule

leads to a (1,4) glycosidic bond or linkage. Similarly, reaction of the anomeric carbon of one D-glucose molecule with the hydroxyl group on carbon atom 6 of a second D-glucose molecule leads to a (1,6) glycosidic bond or linkage. One skilled in the art will recognize that the glycosidic linkages may occur in α or β configurations. Glycosidic linkage configurations are designated by, for example, α(1,4) and α(1,6).

The term "α" refers to the conformation of the linkage being above the plane of the ring. In contrast, a "β" linkage refers to a linkage below the plane of the ring.

The term "oligosaccharide" refers to compounds containing between two and ten monosaccharide units linked by glycosidic linkages. The term "polysaccharide" refers to compounds containing more than ten monosaccharide units linked by glycosidic linkages and generally refers to a mixture of the larger molecular weight species. A polysaccharide composed of a single monomer unit is referred to by the term "homopolysaccharide".

The term "isomaltosaccharide" refers to an oligosaccharide with at least one α(1,6)-linkage.

The term "(1,4) linkage" refers to the relationship of two saccharides in that the C1 from one saccharide unit is bonded to the C4 of the second saccharide unit.

The term "(1,6) linkage" refers to the relationship of two saccharides in that the C1 from one saccharide unit is bonded to the C6 of the second saccharide unit.

The terms "amylase" and "α-amylase" refer to an enzyme that catalyzes the hydrolysis of an α(1,4) glycosidic linkage. The activity, hydrolysis of an α(1,4) glycosidic linkage, is referred to by the terms "amylase activity" or "amylolytic activity". Amylases include but are not limited to the group comprising IUBMB classifications EC 3.2.1.1 (amylase), EC 3.2.1.60 ((1,4)-α- maltotetraohydrolase), and EC 3.2.1.98 ((1,4)-α-maltohexaosidase).

The terms "isoamylase" and "α-isoamylase" refer to an enzyme that catalyzes the hydrolysis of an α(1,6) glycosidic linkage. The activity, hydrolysis of an α(1,6) glycosidic linkage, is referred to by the terms "isoamylase activity" or "isoamylolytic activity". Isoamylases include but are not limited to the group comprising IUBMB classifications EC 3.2.1.10 (oligo-(1,6)-glucosidase), EC 3.2.1.11 (dextranase), EC 3.2.1.41 (pullulanase), and EC 3.2.1.68 (isoamylase).

The terms "glucosidase" and "α-glucosidase" refer to an enzyme that catalyzes the hydrolysis of both an α(1,4) glycosidic linkage and an α(1,6) glycosidic linkage and liberates α-D-glucose units from the non-reducing end of oligosaccharides. A glucosidase has both amylolytic activity and isoamylolytic activity. Glucosidases include but are not limited to the group comprising IUBMB classification EC 3 2.1.3 (amyloglucosidase) and EC 3.2.1.20 (α-Glucosidases).

The term "α(1,6)-linked glucose oligosaccharide hydrolyzing enzyme" refers to an enzyme possessing the functional activity to catalyze the hydrolysis of an α(1,6) glycosidic linkage. Specific examples of an enzyme possessing such a functional activity include isoamylases, α-isoamylases, glucosidases, and α-glucosidases.

The term "isomaltase" or "oligo-(1,6)-glucosidase" or "dextrin 6-α-D-glucanohydrolase" refers to an enzyme (EC 3.2.1.10) that hydrolyzes only α(1,6)-linkages at the nonreducing end of oligosaccharides.

The term "DexB" refers to the (1,6)-α-glucosidase encoded by the dexB gene (GenBank Accession number M77351) of *Streptococcus mutans*, which releases glucose from the non-reducing ends of α(1,6)-linked isomaltosaccharides and dextran.

The term "limit dextrin" refers to the product of the amylolytic degradation of starch comprising monosaccharides and oligosaccharides. The action of amylase on amylopectin yields a mixture of monosaccharide (D-glucose), disaccharides (maltose, α(1,4) linked, and isomaltose, α(1,6) linked) and higher oligosaccharides. The higher oligosaccharides may be linear (contain exclusively α(1,4) linkages) or branched (contain predominantly α(1,4) linkages and α(1,6) linkages).

The term "degree of polymerization" or "DP" refers to the number of monomer units present in an individual component of a saccharide mixture; for example, a monosaccharide such as D-glucose has a DP of 1, a disaccharide such as maltose has a DP of 2, a trisaccharide such as panose has a DP of 3, etc. When applied to polysaccharide mixtures or oligosaccharide mixtures, DP refers to the average number of monomers per molecule.

The term "dextrose equivalent" ("DE") refers to the "reducing sugar content expressed as dextrose percentage on dry matter" as determined by the Lane-Eynon titration. (Handbook of Starch Hydrolysis Products and their Derivatives, M. W. Kearsely and S. Z. Dziedzic, eds., Blackie Academic & Professional, page 86). The DE scale indicates the degree of

hydrolysis of starch, starch having a nominal value of 0 DE and the ultimate hydrolysis product having a value of 100 DE.

Amylase and isoamylase activity may be intracellular or extracellular. For the purposes of this invention, the term "intracellular activity" is meant to refer to enzymatic activity that can be observed with disrupted cells or cell extracts when provided substrate but not with intact cells when provided substrate extracellularly. The term "extracellular activity" is meant to refer to activity that is observed with intact cells (including growing cells) when provided substrate extracellularly. The inability of the enzyme substrates to passively diffuse or be actively transported into the cell is implied in the terms "intracellular activity" and "extracellular activity"

"Target molecule" refers to a biocatalytically-produced product. This may be a compound that is naturally produced by the biocatalyst or non-native genes may be genetically engineered into a microorganism for their functional expression in the biofermentation. "Target molecule" in this context also refers to any by-product of the biofermentation that would be desirable to selectively remove from the biofermentation system to eliminate feedback inhibition and/or to maximize biocatalyst activity.

"Volumetric productivity" refers to the mass of target molecule produced in a biofermentor in a given volume per time, with units of grams/(liter hour) (abbreviated g/(L hr)). This measure is determined by the specific activity of the biocatalyst and the concentration of the biocatalyst. It is calculated from the titer, run time, and the working volume of the biofermentor.

"Titer" refers to the target molecule concentration with units of grams/liter (abbreviated g/L).

The terms "polynucleotide" or "polynucleotide sequence", "oligonucleotide", "nucleic acid sequence", and "nucleic acid fragment" or "isolated nucleic acid fragment" are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof.

The term "isolated" refers to materials, such as nucleic acid molecules and/or proteins, which are substantially free or otherwise

removed from components that normally accompany or interact with the materials in a naturally occurring environment. Isolated polynucleotides may be purified from a host cell in which they naturally occur. Conventional nucleic acid purification methods known to skilled artisans may be used to obtain isolated polynucleotides. The term also embraces recombinant polynucleotides and chemically synthesized polynucleotides.

As used herein, an "isolated nucleic acid molecule" or "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

The term "complementary" is used to describe the relationship between nucleotide bases that are capable to hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, the instant invention also includes isolated nucleic acid fragments that are complementary to the complete sequences as reported in the accompanying Sequence Listing as well as those substantially similar nucleic acid sequences.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof. The terms "substantially similar" and "corresponding substantially" are used interchangeably herein.

Moreover, alterations in a nucleic acid fragment that result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent

product. Nucleotide changes that result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5 % SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5 % SDS at 45 °C for 30 min, and then repeated twice with 0.2X SSC, 0.5 % SDS at 50 °C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5 % SDS was increased to 60 °C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1 % SDS at 65 °C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least 70 % identical, preferably at least 80 % identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are at least 85 % identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least 90 % identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least 95 % identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above identities but typically encode a polypeptide having at least

50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids.

It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying related polypeptide sequences. Useful examples of percent identities are 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, or 95 %, or any integer percentage from 55 % to 100 %. The term " % identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993); Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 60 % identical, preferably at least about 80 % identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are about 85 % identical to the amino acid

sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90 % identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95 % identical to the amino acid sequences reported herein.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also the explanation of the BLAST alogarithm on the world wide web site for the National Center for Biotechnology Information at the National Library of Medicine of the National Institutes of Health). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" or "synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to a nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well-established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of the nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include but is not limited to the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990), and DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 USA), and the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome*

Res., [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, NY). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters that originally load with the software when first initialized.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. A "chimeric protein" is a protein encoded by a chimeric gene. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign-gene" refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, recombinant DNA constructs, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence.

"Regulatory sequences" and "suitable regulatory sequence" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or may be composed of

different elements derived from different promoters found in nature, or may even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters that cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

Promoters which are useful to drive expression of the genes of the present invention in a desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genes is suitable for the present invention including but not limited to: *CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI* (useful for expression in *Saccharomyces*); *AOX1* (useful for expression in *Pichia*); and *lac, ara, tet, trp,* IP_L, IP_R, *T7, tac,* and *trc* (useful for expression in *Escherichia coli*), *Streptomyces lividins* GI, as well as the *amy, apr,* and *npr* promoters and various phage promoters useful for expression in *Bacillus*.

"Translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol.* 3:225-236).

"3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. ((1989) *Plant Cell* 1:671-680).

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA

transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into polypeptides by the cell. "cDNA" refers to DNA that is complementary to and derived from an mRNA template. The cDNA can be single-stranded or converted to double stranded form using, for example, the Klenow fragment of DNA polymerase I. "Sense-RNA" refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to two or more nucleic acid fragments located on a single polynucleotide and associated with each other so that the function of one affects the function of the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

A "protein" or "polypeptide" is a chain of amino acids arranged in a specific order determined by the coding sequence in a polynucleotide encoding the polypeptide. Each protein or polypeptide has a unique function.

"Signal sequence" refers to a nucleotide sequence that encodes a signal peptide.

"Transformation" refers to the transfer of a nucleic acid fragment into a host organism or the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "recombinant", "transgenic" or "transformed" organisms. Thus, isolated polynucleotides of the present invention can be incorporated into recombinant constructs, typically DNA constructs, capable of introduction into and replication in a host cell. Such a construct can be a vector that includes a replication system and sequences that are capable of transcription and translation of a polypeptide-encoding sequence in a given host cell. Typically, expression vectors include, for example, one or more cloned genes under the transcriptional control of 5' and 3' regulatory sequences and a selectable marker. Such vectors also can contain a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally- or developmentally-regulated, or location-specific expression), a transcription initiation start site, a ribosome binding site, a transcription termination site, and/or a polyadenylation signal.

The terms "host cell" or "host organism" refer to a microorganism capable of receiving foreign or heterologous genes or multiple copies of endogenous genes and of expressing those genes to produce an active gene product.

The terms "DNA construct" or "construct" refer to an artificially constructed fragment of DNA. Such construct may be used by alone or may be used in conjunction with a vector.

The terms "plasmid", "vector" and "cassette" refer to an extra chromosomal element often carrying genes that are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of

introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements, in addition to the foreign gene, that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

The terms "encoding" and "coding" refer to the process by which a gene, through the mechanisms of transcription and translation, produces an amino acid sequence. The process of encoding a specific amino acid sequence includes DNA sequences that may involve base changes that do not cause a change in the encoded amino acid, or which involve base changes which may alter one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

"PCR" or "polymerase chain reaction" is well known by those skilled in the art as a technique used for the amplification of specific DNA segments (U.S. Patent Nos. 4,683,195 and 4,800,159).

"ORF" or "open reading frame" is a sequence of nucleotides in a DNA molecule that encodes a peptide or protein. This term is often used when, after the sequence of a DNA fragment has been determined, the function of the encoded protein is not known.

The term "fermentable carbon substrate" refers to a carbon source capable of being metabolized by host organisms of the present invention and particularly those carbon sources selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, and one-carbon substrates or mixtures thereof.

Isolation of Homologs

The nucleic acid fragments of the instant invention may be used to isolate genes encoding homologous proteins from the same or other microbial species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction (PCR), Mullis et al., U.S.

Patent 4,683,202), ligase chain reaction (LCR), Tabor et al., *Proc. Acad. Sci.* USA 82, 1074, (1985)), or strand displacement amplification (SDA, Walker et al., *Proc. Natl. Acad. Sci. U.S.A.*, 89, 392, (1992)).

Typically, in PCR-type amplification techniques, the primers have different sequences and are not complementary to each other. Depending on the desired test conditions, the sequences of the primers should be designed to provide for both efficient and faithful replication of the target nucleic acid. Methods of PCR primer design are common and well known in the art. (Thein and Wallace, "The use of oligonucleotide as specific hybridization probes in the Diagnosis of Genetic Disorders", in *Human Genetic Diseases: A Practical Approach*, K. E. Davis Ed., (1986) pp. 33-50 IRL Press, Herndon, Virginia); Rychlik, W. (1993) In White, B. A. (ed.), Methods in Molecular Biology, Vol. 15, pages 31-39, PCR Protocols: Current Methods and Applications. Humania Press, Inc., Totowa, NJ.)

Hybridization methods are well defined. Typically the probe and sample must be mixed under conditions that will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or organic salt under the proper concentration and temperature conditions. The probe and sample nucleic acids must be in contact for a long enough time that any possible hybridization between the probe and sample nucleic acid may occur. The concentration of probe or target in the mixture will determine the time necessary for hybridization to occur. The higher the probe or target concentration, the shorter the hybridization incubation time needed.

Various hybridization solutions can be employed. Typically, these comprise from about 20 to 60 % volume, preferably 30 %, of a polar organic solvent. A common hybridization solution employs about 30-50 % v/v formamide, about 0.15 to 1M sodium chloride, about 0.05 to 0.1 M buffers, such as sodium citrate, Tris-HCl, PIPES or HEPES (pH range about 6-9), about 0.05 to 0.2 % detergent, such as sodium dodecylsulfate, or between 0.5-20 mM EDTA, FICOLL (Pharmacia Inc.) (about 300-500 kilodaltons), polyvinylpyrrolidone (about 250-500 kDal), and serum albumin. Also included in the typical hybridization solution will be unlabeled carrier nucleic acids from about 0.1 to 5 mg/mL, fragmented nucleic DNA, e.g., calf thymus or salmon sperm DNA, or yeast RNA, and optionally from about 0.5 to 2 % wt./vol. glycine. Other additives may also be included, such as volume exclusion agents that include a variety of polar water-soluble or swellable agents, such as polyethylene glycol,

anionic polymers such as polyacrylate or polymethylacrylate, and anionic saccharidic polymers, such as dextran sulfate.

Recombinant Expression–Microbial

The genes and gene products of the present sequences may be introduced into microbial host cells. Preferred host cells for expression of the instant genes and nucleic acid molecules are microbial hosts that can be found broadly within the fungal or bacterial families and which grow over a wide range of temperature, pH values, and solvent tolerances. Large scale microbial growth and functional gene expression may utilize a wide range of simple or complex carbohydrates, organic acids and alcohols, saturated hydrocarbons such as methane or carbon dioxide in the case of photosynthetic or chemoautotrophic hosts. However, the functional genes may be regulated, repressed or depressed by specific growth conditions, which may include the form and amount of nitrogen, phosphorous, sulfur, oxygen, carbon or any trace micronutrient including small inorganic ions. In addition, the regulation of functional genes may be achieved by the presence or absence of specific regulatory molecules that are added to the culture and are not typically considered nutrient or energy sources. Growth rate may also be an important regulatory factor in gene expression. Examples of suitable host strains include but are not limited to fungal or yeast species such as *Aspergillus*, *Trichoderma, Saccharomyces*, *Pichia*, *Candida*, *Hansenula*, or bacterial species such as member of the proteobacteria and actinomycetes as well as the specific genera *Rhodococcus, Acinetobacter, Arthrobacter, Brevibacterium, Acidovorax, Bacillus*, *Streptomyces*, *Escherichia, Salmonella, Pseudomonas*, and *Cornyebacterium.*

E. coli is particularly well suited to use as the host microorganism in the instant invention fermentative processes. *E. coli* is not able to metabolize oligosaccharides containing an $\alpha(1,6)$ linkage and also has difficulty metabolizing any oligosaccharide of DP > 7.

Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct chimeric genes to produce the any of the gene products of the instant sequences. These chimeric genes could then be introduced into appropriate microorganisms via transformation techniques to provide high-level expression of the enzymes.

Vectors or cassettes useful for the transformation of suitable host cells are well known in the art. Typically the vector or cassette contains sequences directing transcription and translation of the relevant gene, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene harboring transcriptional initiation controls and a region 3' of the DNA fragment which controls transcriptional termination. It is most preferred when both control regions are derived from genes homologous to the transformed host cell, although it is to be understood that such control regions need not be derived from the genes native to the specific species chosen as a production host.

Initiation control regions or promoters, which are useful to drive expression of gene products. Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary, however, it is most preferred if included.

For some applications it will be useful to direct the instant proteins to different cellular compartments. It is thus envisioned that the chimeric genes described above may be further supplemented by altering the coding sequences to encode enzymes with appropriate intracellular targeting sequences such as transit sequences.

Enzymes having enhanced activity

It is contemplated that the present sequences may be used to produce gene products having enhanced or altered activity. Various methods are known for mutating a native gene sequence to produce a gene product with altered or enhanced activity including but not limited to error prone PCR (Melnikov *et al.*, *Nucleic Acids Research*, (Feb. 15, 1999) Vol. 27, No. 4, pp. 1056-1062); site directed mutagenesis (Coombs *et al.*, Proteins (1998), 259-311, 1 plate. Editor(s): Angeletti, Ruth Hogue. Publisher: Academic, San Diego, CA) and "gene shuffling" (US 5,605,793; US 5,811,238; US 5,830,721; and US 5,837,458, incorporated herein by reference).

Pathway Modulation

Knowledge of the sequence of the present genes will be useful in manipulating the sugar metabolism pathways in any organism having such a pathway. Methods of manipulating genetic pathways are common and well known in the art. Selected genes in a particularly pathway may be up-regulated or down-regulated by variety of methods. Additionally,

competing pathways organism may be eliminated or sublimated by gene disruption and similar techniques.

Once a key genetic pathway has been identified and sequenced specific genes may be up-regulated to increase the output of the pathway. For example, additional copies of the targeted genes may be introduced into the host cell on multicopy plasmids such as pBR322. Alternatively the target genes may be modified so as to be under the control of non-native promoters. Where it is desired that a pathway operate at a particular point in a cell cycle or during a fermentation run, regulated or inducible promoters may be used to replace the native promoter of the target gene. Similarly, in some cases the native or endogenous promoter may be modified to increase gene expression. For example, endogenous promoters can be altered *in vivo* by mutation, deletion, and/or substitution (see, Kmiec, U.S. Patent 5,565,350; Zarling et al., PCT/US93/03868).

Within the context of the present invention it may be useful to modulate the expression of the sugar metabolism pathway by any one of a number of well-known methods (e.g., anti-sense, radiation- or chemically-induced mutations, gene-shuffling, etc.). For example, the present invention provides a number of genes encoding key enzymes in the sugar metabolism pathway leading to the production of simple sugars. The isolated genes include the α-glucosidase and isomaltase genes. Where, for example, it is desired to accumulate glucose or maltose, any of the above methods may be employed to overexpress the α-glucosidase and isomaltase genes of the present invention. Similarly, biosynthetic genes' accumulation of glucose or maltose may be effected by the disruption of down stream genes such as those of the glycolytic pathway by any one of the methods described above.

Biofermentations

The present invention is adaptable to a variety of biofermentation methodologies, especially those suitable for large-scale industrial processes. The invention may be practiced using batch, fed-batch, or continuous processes, but is preferably practiced in fed-batch mode. These methods of biofermentation are common and well known in the art (Brock, T. D.; *Biotechnology: A Textbook of Industrial Microbiology*, 2nd ed.; Sinauer Associates: Sunderland, MA (1989); or Deshpande, *Appl. Biochem. Biotechnol.* 36:227 (1992)).

"Biofermentation system" or "biofermentation" refers to a system that uses a biocatalyst to catalyze a reaction between substrate(s) and product(s).

The Biocatalyst

The biocatalyst initiates or modifies the rate of a chemical reaction between substrate(s) and product(s). The biocatalyst may be whole microorganisms or in the form of isolated enzyme catalysts. Whole microbial cells can be used as a biocatalyst without any pretreatment such as permeabilization. Alternatively, the whole cells may be permeabilized by methods familiar to those skilled in the art (e.g., treatment with toluene, detergents, or freeze-thawing) to improve the rate of diffusion of materials into and out of the cells.

Microorganisms useful in the present invention may include, but are not limited to, bacteria (such as the enteric bacteria *Escherichia* and *Salmonella*, for example, as well as *Bacillus*, *Acinetobacter*, *Streptomyces*, *Methylobacter*, *Rhodococcus,* and *Pseudomonas*); cyanobacteria (such as *Rhodobacter* and *Synechocystis*); yeasts (such as *Saccharomyces*, *Zygosaccharomyces*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Pichia,* and *Torulopsis*); filamentous fungi (such as *Aspergillus* and *Arthrobotrys*); and algae. For purposes of this application, "microorganism" also encompasses cells from insects, animals, or plants.

Culture Conditions

Materials and methods suitable for maintenance and growth of microbial cultures are well known to those in the art of microbiology or biofermentation science art (Bailey and Ollis, *Biochemical Engineering Fundamentals*, 2nd Edition; McGraw-Hill: NY (1986)). Consideration must be given to appropriate media, pH, temperature, and requirements for aerobic, microaerobic, or anaerobic conditions, depending on the specific requirements of the microorganism for the desired functional gene expression.

Media and Carbon Substrates

Biofermentation media (liquid broth or solution) for use in the present invention must contain suitable carbon substrates, chosen in light of the needs of the biocatalyst. Suitable substrates may include, but are not limited to, monosaccharides (such as glucose and fructose), disaccharides (such as lactose or sucrose), oligosaccharides and polysaccharides (such as starch or cellulose or mixtures thereof), or unpurified mixtures from renewable feedstocks (such as cheese whey

permeate, cornsteep liquor, sugar beet molasses, and barley malt). The carbon substrate may also be one-carbon substrates (such as carbon dioxide, methanol, or methane).

In addition to an appropriate carbon source, biofermentation media must contain suitable minerals, salts, vitamins, cofactors, buffers, and other components, known to those skilled in the art (Bailey and Ollis, *Biochemical Engineering Fundamentals*, 2nd ed; pp 383-384 and 620-622; McGraw-Hill: New York (1986)). These supplements must be suitable for the growth of the biocatalyst and promote the enzymatic pathway necessary to produce the biofermentation target product.

Finally, functional genes that express an industrially useful product may be regulated, repressed, or derepressed by specific growth conditions (for example, the form and amount of nitrogen, phosphorous, sulfur, oxygen, carbon or any trace micronutrient including small inorganic ions). The regulation of functional genes may be achieved by the presence or absence of specific regulatory molecules (such as gratuitous inducers) that are added to the culture and are not typically considered nutrient or energy sources. Growth rate may also be an important regulatory factor in gene expression.

EXAMPLES

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "mL" means milliliter(s), "L" means liter(s),"mM" means millimolar, "nm" means nanometer, "g" means gram(s), and "kg" means kilogram(s), "HPLC" means high performance liquid chromatography, "RI" means refractive index.

GENERAL METHODS:

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in *Manual of Methods for*

General Bacteriology; Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, Eds., American Society for Microbiology: Washington, D.C. (1994) or in *Biotechnology: A Textbook of Industrial Microbiology*; Brock, T. D., 2nd ed.; Sinauer Associates: Sunderland, MA (1989).

The conversion of glycerol to 1,3-propanediol was monitored by HPLC. Analyses were performed using standard techniques and materials available to one of skill in the art of chromatography. One suitable method utilized a Waters Maxima 820 HPLC system using UV (210 nm) and RI detection. Samples were injected onto a Shodex SH-1011 column (8 mm x 300 mm, purchased from Waters, Milford, MA) equipped with a Shodex SH-1011P precolumn (6 mm x 50 mm), temperature controlled at 50 °C, using 0.01 N H_2SO_4 as mobile phase at a flow rate of 0.5 mL/min. When quantitative analysis was desired, samples were prepared with a known amount of trimethylacetic acid as external standard. Typically, the retention times of glucose (RI detection), glycerol, 1,3-propanediol (RI detection), and trimethylacetic acid (UV and RI detection) were 15.27 min, 20.67 min, 26.08 min, and 35.03 min, respectively.

Example 1

Genome Sequencing of *Bifidobacterium breve ATCC* 15700

Bifidobacterium breve (ATCC 15700) was purchased from the American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108, U.S.A. A cell pellet was obtained and suspended in a solution containing 10 mM Na-EDTA and 50 mM Tris-HCl, pH 7.5. Genomic DNA was isolated from *Bifidobacterium breve* (ATCC 15700) according to standard protocols. Genomic DNA and library construction were prepared according to published protocols (Fraser et al., *Science* 270 (5235):397-403 (1995)).

Genomic DNA preparation: After suspension, the cells were gently lysed in 0.2 % sarcosine, 20 mM beta-mercaptoethanol, and 150 units/mL of Lyticase and incubated for 30 min at 37 °C. DNA was extracted twice with Tris-equilibrated phenol and twice with chloroform. DNA was precipitated in 70 % ethanol and suspended in a solution containing 1 mM Na-EDTA and 10 mM Tris-HCl, pH 7.5. The DNA solution was treated with a mix of RNAases, then extracted twice with Tris-equilibrated phenol and twice with chloroform. This was followed by precipitation in ethanol and suspension in 1 mM Na-EDTA and 10 mM Tris-HCl, pH 7.5.

Library construction: 50 to 100 μg of chromosomal DNA was suspended in a solution containing 30 % glycerol, 300 mM sodium

acetate, 1 mM Na-EDTA, and10 mM Tris-HCl, pH 7.5 and sheared at 12 psi for 60 sec in an Aeromist Downdraft Nebulizer chamber (IBI Medical products, Chicago, IL). The DNA was precipitated, suspended and treated with BAL-31 nuclease. After size fractionation on a low melt agarose gel, a fraction (2.0 kb or 5.0 kb) was excised, cleaned, and ligated to the phosphatased *Sma*I site of pUC18 (Amersham Biosciences) using T4 DNA ligase (New England Biolabs, Inc., Beverly, MA). The ligation mix was run on a gel and the DNA band representing the vector plus one insert ligation product was excised, treated with T4 DNA polymerase (New England Biolabs), and then religated. This two-step ligation procedure was applied to produce a high titer library with greater than 99 % single inserts.

Sequencing: A shotgun sequencing strategy approach was adopted for the sequencing of the whole microbial genome (Fleischmann, R. et al., *Science* 269(5223):496-512 (1995)). Sequence was generated on an ABI Automatic sequencer (Applied Biosystems, Foster City, CA) using dye terminator technology (U.S. 5,366,860; EP 272,007) using a combination of vector and insert-specific primers. Sequence editing was performed in either DNAStar (DNA Star Inc., Madison, WI) or the Wisconsin GCG program (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI) and the CONSED package (version 7.0). All sequences represent coverage at least two times in both directions. Sequence assembly was performed using the Phred/Phrap software package (version 0.961028.m / 0.990319).

EXAMPLE 2

Identification of Carbohydrate Degradation Genes

Genes encoding isoamylase activity were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al., *J. Mol. Biol.* 215:403-410 (1993); see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The sequences obtained were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences

contained in the "nr" database using the BLASTP algorithm (Gish and States, *Nature Genetics 3*:266-272 (1993)) provided by the NCBI.

All comparisons were done using either the BLASTN or BLASTP algorithm. The results of the BLAST comparison are presented in Table 1, which summarizes the sequences to which they have the most similarity. Table 1 displays data based on the BLASTP algorithm with values reported in expectation values. The expectation value (E-value) is the number of different alignments with scores equivalent to or better than a particular score S that are expected to occur in a database search by chance. The lower the E-value, the more significant the score.

TABLE 1

Clone Name	Similarity Identified	SEQ ID	SEQ ID	% Identity[a]	% Similarity[b]	E-value[c]	Citation
mbc1g.pk007.h12 *Bifidobacterium breve*	(AF411186) alpha-glucosidase [*Bifidobacterium adolescentis*]	1	2	61	70	0.0	Van den Broek, L.A.M. et al. "Cloning and characterization of two alpha-glucosidases from *Bifidobacterium adolescentis*" NCBI database
Mbc2g.pk018.j20 *Bifidobacterium breve*	(AF358444) alpha-glucosidase [*Bifidobacterium adolescentis*]	3	4	73	84	0.0	Van den Broek, L.A.M. et al. "Cloning and characterization of two alpha-glucosidases from *Bifidobacterium adolescentis*" NCBI database
mbc1g.pk026.k1 *Bifidobacterium breve*	(AF358444) alpha-glucosidase [*Bifidobacterium adolescentis*]	5	6	69	82	0.0	Van den Broek, L.A.M. et al. "Cloning and characterization of two alpha-glucosidases from *Bifidobacterium adolescentis*" NCBI database
DexB *Steptococcus mutans*	(M77351) dextran glucosidase [*Streptococcus mutans*]	16	17	100	100	0.0	Russell, R.R. and Ferretti, J.J. "Nucleotide sequence of the dextran glucosidase (*dexB*) gene of *Streptococcus mutans* " *J. Gen. Microbiol.* 136 (Pt 5), 803-810 (1990)

[a] %Identity is defined as percentage of amino acids that are identical between the two proteins.

[b] % Similarity is defined as percentage of amino acids that are identical or conserved between the two proteins.

[c] Expectation value. The Expectation value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

EXAMPLE 3

Intracellular isoamylase activity in *E. coli* containing the *Streptococcus mutans dexB* gene

For cloning of the *dexB* gene, genomic DNA was isolated from *Streptococcus mutans* (ATCC 25175D) using the protocol described in Jagusztyn et al. (*J. Gen. Microbiol.* 128:1135-1145(1982)).

Oligonucleotide primers (SEQ ID NO:7 and SEQ ID NO:8) were designed based on *Streptococcus mutans* (*dexB*) DNA sequence (Ferretti et al., *Infection and Immunity* 56:1585-1588 (1988)) and also included *Bam*HI and *Sal*I restriction sites. The *dexB* gene was amplified using the standard PCR protocol included with the HotStartTaq™ kit (Qiagen, Valencia, CA). Reactions contained 1 ng of genomic DNA and 1 μM each of primers. The resulting 1.6 kb DNA fragment was digested with the enzymes *Bam*HI and *Sal*I. The digested fragment was cloned directly into the plasmid pTRC99a (amp^R) (Amersham-Pharmacia, Amersham, UK) resulting in a translational fusion with the LacZ gene. The plasmid, designated pTRC99-dexB, also contains the coding sequence for the first 10 amino acids of the LacZ gene, which upon expression are fused to the N-terminal end of native DexB protein. pTRC99-dexB plasmid was transformed into *E. coli* DH5α cells using the manufacturer's protocol (Invitrogen, Carlsbad, CA) and plated on Luria Broth (LB) medium containing 100 μg/mL ampicillin.

Isoamylase activity was assessed from crude protein extract following expression in *E. coli*. A single colony of *E. coli* DH5α/pTRC99-dexB was cultured overnight in LB medium and then diluted 1:100 into fresh LB medium (3.0 mL) and cultured for an additional two hr at 37°C. Following this incubation, the DexB gene was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Crude protein was extracted from induced cells following an additional two hr incubation. To isolate the crude protein extract, cells were collected by centrifugation (1 x 8000 g) and then suspended in 0.5 mL of phosphate buffer (10 mM, pH 6.8). The suspension was sonicated to release total cellular protein and centrifuged (1 x 14,000 g) to remove cell debris. Total protein present in the supernatant was assayed for isoamylase activity by incubation with isomaltose or separately with panose at 37 °C in 10 mM phosphate buffer (pH 6.8) for two hrs. Products of the reaction were characterized by High Performance Anion Exchange Chromatography (HPAEC).

For HPAEC, samples were prepared and analyzed in the following manner. After the two-hr incubation with isomaltose or panose, total protein extracts were filtered through a 0.22 μM Spin-X (R) centrifuge tube filter (Costar, Corning, NY) and diluted with sterile filtered water. Samples were analyzed by HPAEC (Dionex, Sunnyvale, CA) using a PA10 column, 100 mM sodium hydroxide as the eluent and a 0-150 mM sodium acetate linear gradient. Results demonstrating degradation of isomaltose using pTRC99-dexB cell-extract are listed in Table 2. Degradation of panose, and the products formed by incubation with pTRC99-dexB cell-extract are listed in Table 3.

Table 2

Activity of DexB Crude Protein Extract with Isomaltose (250 μg/mL)

Cell Line	Isomaltose (μg/mL)
DH5α/pTRC99a (negative control)	256
DH5α/pTRC99-dexB	ND

ND = not detected

Table 3

Activity of DexB Crude Protein Extracts with Panose (150 μg/mL)

Cell Line	Panose (μg/mL)	Maltose (μg/mL)	Isomaltose (μg/mL)	Glucose (μg/Ml)
DH5α/pTRC99a (negative control)	122	ND	ND	ND
DH5α/pTRC99-dexB	ND	74	8	82

ND = not detected

EXAMPLE 4

Expression of the *Bifidobacterium breve* isoamylolytic genes in *E. coli*

Several open reading frames from the *Bifidobacterium breve* (ATCC 15700) library were identified as putative candidate genes with activity against α(1,6)-linked glucose oliogosaccharides (Example 2). Three putative clones, mbc1g.pk007.h12 (h12), mbc1g.pk026.k1 (k1), and mbc2g.pk018.j20 (j20) were chosen for detailed characterization of

isoamylolytic activity, using oliogosaccharides containing α(1,6)-linked glucose.

E. coli DH5α strains containing the cloned full length coding sequence of the putative isoamylolytic *Bifidobacterium* genes in pUC18 from Example 1 were inoculated to LB medium and cultured at 37°C. The culture was diluted after 20 hr (1:100) in fresh LB medium and incubated for an additional 3-4 hr at 37 °C. Total protein extract was prepared from cells as described in Example 3. Total protein present in the supernatant was assayed for isoamylolytic activity by incubation with isomaltose or separately with panose at 37 °C in 10 mM phosphate buffer (pH 6.8) for two hr. Samples were prepared and products of the reaction were characterized by High Performance Anion Exchange Chromatography (HPAEC) as described in Example 3. Results demonstrated that the enzymes produced from clones h12, k1, and j20 degraded isomaltose to glucose (Table 4).

Table 4

Activity of *B. breve* crude extracts with Isomaltose (150 μg/mL)

Cell line	Isomaltose (μg/mL)	Glucose (μg/mL)
DH5α/pUC18 (negative control)	107	37
DH5α – h12	6	187
DH5α – k1	5	165
DH5α – j20	8	154

ND = not detected

Total protein extracts were incubated with panose (250 μg/mL) for two hr and then filtered through a 0.22μM Spin-X (R) centrifuge tube filter (Costar, Corning, NY). Samples were analyzed by HPAEC as described in Example 3. The absence of panose following incubation demonstrated that the enzymes produced from the clones h12, k1 and j20 are capable of degrading panose. Figure 1 shows that the clone h12 degrades panose completely to glucose (also shown is the negative control, plasmid pUC18 in *E. coli* DH5α). Figure 1 also shows that the enzymes from the k1 and j20 clones degrade panose to glucose and maltose.

EXAMPLE 5

Expression of the native *B. breve* j20 isoamylase gene in *E. coli*

The native *Bifidobacterium breve* gene j20 (obtained in Example 1) appeared to have a signal peptide at the NH-end of the mature coding sequence (determined by pSort prediction software; Nakai and Kanehisa, Expert, PROTEINS: *Structure, Function, and Genetics* 11:95-110 (1991)). The nucleic and amino acid sequences for the *Bifidobacterium breve* j20 gene, which codes for an α(1,6)-linked glucose oligosaccharide hydrolyzing activity, are SEQ ID NO:30 SEQ ID NO:31, respectively.

Metabolism of isomaltose was, therefore, attempted using intact whole cells. This was accomplished by culturing a single colony of *E. coli* DH5α cells expressing the j20 gene in LB medium containing isomaltose (500 μg/mL) at 37 °C for 24 hr. Following incubation, cells were removed from the medium, and the medium was prepared and analyzed by HPAEC methods described in Example 3. The presence of extracellular isoamylase activity in cells expressing the *B. breve* j20 gene was demonstrated by reduced levels of isomaltose compared to the negative control (*E. coli* DH5α cells containing only the original pUC18 plasmid). The results in Table 5 demonstrate that *E. coli* cells expressing the native j20 gene degraded isomaltose supplied extracellularly.

Table 5

Isomaltose Metabolized by the Native j20 Gene

Cell line	Isomaltose (μg/mL)	Glucose (μg/mL)
DH5α/pUC18 (negative control)	508	26
DH5α – j20	180	22

EXAMPLE 6

Extracellular targeting of the *S. mutans* dexB and *B. breve* isoamylase enzymes

Because the *Bifidobacterium breve* k1 and *Streptococcus mutans* *dexB* genes do not appear to contain native signal peptides (pSort prediction software; Nakai and Kanehisa, Expert, PROTEINS: *Structure, Function, and Genetics* 11:95-110 (1991)), the mature coding sequences

were linked in a translational fusion to signal peptides by PCR methods, allowing extracellular expression.

Modular expression vectors containing the *Bacillus subtilis* alkaline and neutral protease genes were constructed in a series of steps beginning with the plasmids pBE505 (Borchert and Nagarajan, *J. Bacteriol.* 173:276-282 (1991)) and pBE311 (Nagarajan and Borchert, *Res. Microbiol.* 142:787-792 (1991)). The plasmids were digested with the restriction enzymes *Kpn*I and *Nru*I. The resulting 969 bp *Kpn*I-*Nru*I fragment from pBE505 was isolated and ligated into the large 7.2 kb *Kpn*I-*Nru*I fragment from pBE311, resulting in pBE559.

Plasmids pBE559 and pBE597 (Chen and Nagarajan, *J. Bacteriol.* 175:5697-5700 (1993)) were then digested with the restriction enzymes *Kpn*I and *Eco*RV. The 941bp *Kpn*I-*Eco*RV fragment from pBE559 was ligated into the 8.9kb *Kpn*I-*Eco*RV fragment from pBE597, resulting in plasmid pBE592.

Plasmid pBE26 (Ribbe and Nagarajan, *Mol. Gen. Genet.* 235:333-339 (1992)) was used as a template to amplify the *B. amyloliquefaciens* alkaline protease (apr) promoter region using PCR methods described in Example 3. The oligonucleotide primer SEQ ID NO:9 was designed and synthesized to introduce an *Nhe*I restriction site at the alkaline protease signal cleavage site and an *Eco*RV restriction site immediately downstream of the cleavage site. The oligonucleotide primer SEQ ID NO:10 was designed to anneal to the 5' polylinker region upstream of the apr promoter region in pBE26. A PCR reaction was carried out using the described primers and plasmid pBE26 template DNA. The resulting 1.2 kb PCR product was digested with *Kpn*I and *Eco*RV and ligated into the large *Kpn*I-*EcoR*V fragment from pBE592, resulting in pBE92.

Plasmid pBE80 (Nagarajan et al., *Gene* 114:121-126 (1992)) was used as a template to amplify the *B. amyloliquefaciens* neutral protease (npr) promoter region using PCR methods described in Example 3. The downstream primer SEQ ID NO:11 was designed and synthesized to introduce an *Nhe*I restriction site at the neutral protease signal cleavage site and an *Eco*RV restriction site immediately downstream of the cleavage site. The primer SEQ ID NO:12 was designed to anneal to the 5' region of the Npr promoter in pBE80. A PCR reaction was carried out using the described primers and DNA template. The resulting 350 bp PCR

product was enzymatically digested with *Kpn*I and *Eco*RV and ligated into the large *Kpn*I-*Eco*RV fragment from pBE592, resulting in pBE93.

A translational fusion of the k1 and dexB genes to signal peptides of the *Bacillus subtilis* alkaline and neutral protease genes in the vectors pBE92 and pBE93 was accomplished using oligonucleotide primers described in Table 6. PCR amplfication was performed by the protocol described in Example 3, using genomic DNA from *Bifidobacterium breve* (ATCC 15700) or pTRC99-dexB plasmid, respectively, as a template.

Oligonucleotide primers SEQ ID NO:14 and SEQ ID NO:15, engineered with *Nhe*I and *Bam*HI sites, were used to amplify a 1.8 kb k1 gene DNA fragment. Oligonucleotide primers SEQ ID NO:13 and SEQ ID NO:8, containing *Nhe*I and *Sal*I restriction enzyme sites, resulted in amplification of a 1.6 kb *dexB* gene DNA fragment. The fragments were digested with the appropriate enzymes and cloned into modular vectors pBE92 and pBE93.

The resulting plasmids (designated pBE92-dexB, pBE93-dexB, pBE92-k1, and pBE93-k1, respectively) contained the native enzyme linked in a translational fusion to the signal peptide such that the signal peptide cleavage site (Ala Ser Ala) was conserved. Nucleic and amino acid sequences for the *Bacillus subtilis* neutral protease signal peptide linked to the *Bifidobacterium breve* k1 gene are SEQ ID NOs:40 and 41, respectively. Nucleic and amino acid sequences for the *Bacillus subtilis* neutral protease signal peptide linked to the *Streptococcus mutans* dexB gene are SEQ ID NOs:42 and 43, respectively. The plasmids were transformed into *E. coli* DH5α cells using the manufacturer's protocol (Invitrogen, Carlsbad, CA) and plated on Luria Broth (LB) medium containing ampicillin (100 μg/mL).

Characterization of activity in *E. coli* DH5α cells containing the pBE93 (negative control), pBE93-dexB or pBE93-k1 plasmid was carried out by inoculating 3.0 mL of LB medium containing ampicillin (100 μg/mL) and isomaltose (0.250 mg/mL). The cells were grown at 37°C for 20 hr. Following incubation, cells were removed from the medium and prepared and analyzed by methods described in Example 3. The presence of extracellular isoamylase activity in cells containing the pBE93, pBE93-dexB or pBE93-k1 plasmid was demonstrated by reduced levels of isomaltose compared to the negative control (*E. coli* DH5α cells containing only the original pBE92 plasmid). The results in Table 6 demonstrate that the Npr-gene fusion proteins degraded isomaltose supplied extracellularly.

Table 6

DexB and K1 Extracellular Fusion Protein Activity in *E. coli* DH5α cells

Cell line	Isomaltose (μg/mL)
pBE93 (negative control)	215
pBE93-dexB (isolate 4)	117
pBE93-dexB (isolate 8)	89
pBE93-k1 (isolate 7)	76
pBE93-k1 (isolate 8)	74
pBE93-k1 (isolate 9)	62

E. coli DH5α cells containing the pBE93-dexB or pBE93-k1 plasmids degraded isomaltose; however, cell growth in minimal media containing isomaltose as the sole carbon source is a much more stringent measure of isoamylase activity. Therefore pBE93-dexB and pBE93-k1 plasmids were transformed into the *E. coli* strain FM5. The FM5 strain, unlike DH5α, has the ability to grow in a minimal medium, containing only salts and trace metals in addition to a carbon source (Maniatis et al. (1982) Molecular Cloning; a Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y; Neidhardt (1987) *Escherichia coli* and *Salmonella typhimurium*, ASM Press, Washington, DC). Native FM5 cells, like the DH5α strain, cannot utilize isomaltose as a carbon source. To confirm this, FM5 cells transformed with the plasmid pBE93 were inoculated into M9 media (Maniatis et al., *supra*; Neidhardt, *supra*) containing either glucose (1 mg/mL) or isomaltose (1 mg/mL) and incubated at 37 °C for at least 20 hr. Cell growth was observed after 20 hr in flasks containing glucose, but not in flasks containing isomaltose, even after a 60 hr incubation.

In contrast to the negative control, FM5 cells containing the Npr-DexB and Npr-k1 fusion proteins (pBE93-dexB and pBE93-k1, respectively) grew well in M9 medium containing isomaltose following a 20 hr incubation period. For this experiment FM5/pBE93, FM5/pBE93-dexB and FM5/pBE93-k1 strains were inoculated into 2.0 mL M9 medium supplemented with either glucose or isomaltose (1 mg/mL) as the sole carbon source. The results, shown in Table 7, indicated that when the dexB or k1 genes, are linked in a translational fusion to the Npr signal

peptide, are expressed in FM5 cells, isomaltose is metabolized and supports cell growth.

Table 7

DexB and K1 Extracellular Fusion Protein Activity in *E. coli* FM5 cells

Cell line	Isomaltose (μg/mL)
pBE93 (negative control)	1091
pBE93-dexB (isolate 2)	319
pBE93-dexB (isolate 15)	197
pBE93-dexB (isolate 3)	183
pBE93-k1 (isolate 5)	34
pBE93-k1 (isolate 4)	20
pBE93-k1 (isolate 3)	17

EXAMPLE 7

Expression of the Npr-dexB and Npr–k1 fusion genes in *E. coli* leads to increased synthesis of various fermentation products

The ability of production hosts to metabolize oligosaccharides containing α(1,6)-linked glucose residues may increase the yield of a fermentation product when a mixture of sugars is supplied as the carbon source. The ability of the Npr-dexB and Npr-k1 fusion proteins to degrade α(1,6)-linkages was tested by first transforming the plasmids pBE93-dexB and pBE93-k1 into a cell line engineered to produce glycerol.

One microgram of plasmid DNA was used to transform *E. coli* strain RJ8n (ATCC PTA-4216), which also contained the plasmid pSYCO101 ($spec^R$) (described in U.S. Patent Application 10/420,587 herein incorporated by reference), which encodes the *DAR1* and *GPP2* genes from *Saccharomyces cerevisiase*, and *dhaB* and *orf* operons from *Klebsiella pnuemoniae*. The transformed *E. coli* strain produces glycerol from glucose as well as 1,3-propanediol when vitamin B12 is added. Methods for the production of glycerol and 1,3-propanediol from glucose are described in detail in U.S. Patent No. 6,358,716 and U.S. Patent No. 6,013,494 herein incorporated by reference. The transformed RJ8n cells were plated on LB medium containing 50 μg/mL spectinomycin and 100 μg/mL ampicillin. Single colonies were used to inoculate 2.0 mL of TM2 medium (potassium phosphate, 7.5 g/L; citric acid, 2.0 g/L;

ammonium sulfate, 3.0 g/L; magnesium sulfate, 2.0 g/L; calcium chloride, 0.2 g/L; ferric ammonium citrate, 0.33 g/L; yeast extract (Difco-BD, Sparks, MD) 5.0 g/L; trace elements (zinc sulfate, copper sulfate, cobalt chloride, manganese sulfate, ferric sulfate, sodium chloride); ammonium hydroxide, pH to 6.5; also containing glucose or isomaltose (1 mg/mL). Cultures were grown for 24 hr at 37 °C. Cells were prepared and analyzed by methods described in Example 3.

Glycerol was shown to accumulate when *E. coli* RJ8n cells containing only the plasmid pSYCO101 were cultured for 24 hr at 37°C in TM2 medium with glucose as the carbon source (Table 8). However, this negative control line produced negligible levels of glycerol when isomaltose was substituted for glucose in the medium, demonstrating that α(1,6)-linked glucose does not support accumulation of a fermentation product. By contrast, glycerol was produced in *E. coli* RJ8n containing the plasmids pSYCO101 and pBE93-dexB or pBE93-k1 when either isomaltose or glucose was provided as sole carbon sources (Table 8). When isomaltose was used as a carbon source, glycerol production was shown to be 8 to 9 times higher in *E. coli* RJ8n containing both the pBE93-dexB and pSYCO101 plasmids as compared to the negative control line, RJ8n containing only pSYCO101. Glycerol accumulation, using isomaltose, was 6 to 10 times higher in lines containing pSYCO101 and pBE93-k1 as compared to the negative control. The data in Table 8 demonstrate that expression of the Npr-dexB or Npr-k1 genes resulted in glycerol production in cultures supplied with isomaltose. The data further demonstrate that levels of product accumulated were comparable for cultures containing the fusion proteins regardless of whether the carbon source was glucose or isomaltose.

Table 8

Glycerol Accumulation Due to Expression of Npr-DexB or Npr-K1

Cell line	Glycerol Accumulated (μg/mL)	
	Glucose-supplied cultures	Isomaltose-supplied cultures
RJ8n/pSYCO101	430	39
RJ8n/pSYCO101/pBE93-dexB (isolate 4)	381	362
RJ8n/pSYCO101/pBE93-dexB (isolate 8)	353	354
RJ8n/pSYCO101/pBE93-k1 (isolate 2)	383	401
RJ8n/pSYCO101/pBE93-k1 (isolate 6)	412	226

The capability of *E. coli* line RJ8n containing the plasmids pSYCO101 and pBE93-k1 to produce fermentation products using α(1,6)-linked glucose as a substrate was further characterized by culturing in TM2 medium containing panose (1 mg/mL) and comparing the results to the same line using glucose as a substrate (1 mg/mL).

Data in Table 9 also show that *E. coli* strain RJ8n containing only the plasmid pSYCO101 (negative control) does not synthesize glycerol when panose is supplied as the sole carbohydrate source in TM2 medium. However, glycerol is produced when the plasmid pBE93-k1 is present in this same strain and cultured in TM2 medium with panose. Glycerol accumulation in *E. coli* RJ8n containing the plasmids pSYCO101 and pBE93-k1 was comparable when either glucose or panose was supplied as a carbohydrate source.

Table 9

Glycerol Accumulation Due to Expression of Npr-K1

Cell line	Glycerol Accumulated (μg/mL)	
	Glucose-supplied cultures	Isomaltose-supplied cultures
RJ8n/pSYCO101	417	25
RJ8n/pSYCO101/pBE93-k1 (9)	396	363
RJ8n/pSYCO101/pBE93-k1 (7)	376	347

The data above demonstrate that expression of the Npr-dexB or Npr-k1 fusion protein in *E. coli* results in increased production of glycerol when isomaltose or panose represents the sole carbohydrate source in the medium. Demonstrating that this result is not limited to glycerol production alone was accomplished by synthesis of another fermentation product (1,3-propanediol) using the same fusion protein expression system.

RJ8n cells transformed with the plasmids pSYCO101 and pBE93-dexB or pBE93-k1 were used to inoculate 2.0 mL of TM2 medium (described above) also containing glucose (1 mg/mL) or isomaltose (1 mg/mL) and vitamin B12 (100 ng/L). Cultures were grown for 20 hr at 37 °C. Cells were prepared and analyzed by methods described in Example 3.

The data in Table 10 demonstrate that 1,3-propanediol was not synthesized by the negative control line (RJ8n/pSYCO101) when grown in media containing only isomaltose as a carbohydrate source. However, when either the Npr-dexB or Npr-k1 fusion protein was expressed in RJ8n cells, isomaltose was shown to be metabolized. This resulted in accumulation of the fermentation product 1,3-propanediol. The data further demonstrate that the level of 1,3-propanediol synthesized by RJ8n cells expressing the Npr-dexB or Npr-K1 fusion protein was comparable whether glucose or isomaltose was supplied as the sole carbohydrate.

Table 10

1,3-Propanediol Accumulation Due to Expression of Npr--dexB or Npr-k1

Cell line	1,3-Propanediol (mg/mL)		
	Glucose-supplied cultures	Isomaltose-supplied cultures	Isomaltose (μg/mL)
RJ8n/pSYCO101	2.8	ND	1225
RJ8n/pSYCO101/pBE93-k1 (9)	1.7	2.8	12
RJ8n/pSYCO101/pBE93-k1 (7)	3.0	2.9	14
RJ8n/pSYCO101/pBE93-dexB	3.0	3.1	27

ND = not detected

EXAMPLE 8

Expression of the *B. breve* k1 gene in *E. coli* using an alternative promoter

The use of alternative promoters to direct expression of a preferred gene is often highly desirable. Alternative promoters may be used to vary the level or timing of gene expression and, therefore, increase utilization of a preferred substrate.

Effective expression of the *B. breve* k1 isoamylase gene using an alternative promoter was demonstrated by replacing the neutral protease promoter in the plasmid pBE93-k1 (Example 6) with a glucose isomerase (GI) promoter and variant of the GI-promoter. Isolation of the *Streptomyces lividins* GI-promoter and creation of the variant promoter was disciosed in U.S. Patent Application 10/420,587. Prior to replacing the NPR-promoter, modifications of the non-coding nucleotide sequences of the neutral protease signal peptide and K1 gene were made. The sequence modifications resulted in restriction enzyme sites, which would be used in subsequent cloning steps.

The restriction enzyme sites SacI and PacI were added to the 5' and 3'-ends of the neutral protease signal peptide and K1 gene sequences, respectively, by PCR using the primers SEQ ID NO. 18 and SEQ ID NO. 19. PCR amplfication was performed by the protocol described in Example 3. A 1919 bp PCR product was isolated and ligated into the pSYCO109mcs wild-type GI *yqhD* plasmid as disclosed in U.S. Patent Application 10/420,587, which was also digested with the enzymes SacI and PacI. The resulting plasmid contains a wild-type GI promoter and the NPR-signal sequence linked in a translational fusion to the k1 gene. This construct was designated WTGI-ss-K1. A variant GI promoter was also used to direct expression of the NPR-signal peptide/K1 fusion. A 1919 bp PCR product, resulting from a reaction using the primers SEQ ID NO:18 and SEQ ID NO:19 was placed into the pSYCO109mcs-short 1.6 GI *yqhD* plasmid, using SacI and PacI restriction enzyme sites. The resulting plasmid was designated LowGI-ss-K1. This variant promoter when operably linked to a yqhD gene was previously shown to direct lower levels of gene expression (U.S. Patent Application 10/420,587) as compared to the wild-type GI promoter-yqhD construct.

Demonstrating effective expression of the K1 gene using the wild-type and variant GI promoters was accomplished by an activity assay. *E. coli* cells (strain DH5α, Invitrogen, Carlsbad, CA) were transformed with the plasmids WTGI-ss-K1 and LowGI-ss-K1 and grown overnight in LB medium. Cell pellets were recovered by centrifugation and suspended in 1/10 volume sodium-phosphate buffer (10 mM, pH 7.0). The cells in the suspension were lysed with a French press and cell-debris was removed by centrifugation. Total protein concentration was determined by Bradford assay (Bio-Rad, Hercules, CA). Activity of the K1 gene product in a total protein isolate was assayed using 4-nitrophenyl-α-D-glucopyranoside

(PNPG, Sigma, ST. Louis, MO). Total protein extract from cells containing the plasmids WTGI-ss-K1, LowGI-ss-K1, NPR-ss-K1 (positive control) and pSYCO109 (negative control) were incubated in a10 mM sodium phosphate buffered solution containing 10 mM PNPG for up to 30 min at 30 °C. Release of the glucose residue from PNPG results in PNP accumulation, which absorbs light at 400 nm. PNP accumulation as a direct result of k1 enzyme activity was monitored over time by absorbance at a wavelength of 400 nm. Table 11 below demonstrates that a promoter, other than the neutral protease promoter, may be used to direct expression of an active K1 gene. The results also demonstrate that an alternative promoter may be used to modify the level of K1 expression and that K1 activity corresponds to the relative level of promoter strength.

Table 11

Rate of PNP production resulting from K1 enzyme activity

Plasmid	Activity (mM PNP/mg protein min^{-1})
WTGI-ss-K1 (high expresser)	0.0144
NPR-ss-K1 (positive control)	0.0104
LowGI-ss-K1 (low expresser)	0.0028
pSYCO109 (negative control)	0.0002

EXAMPLE 9

Integration of the *B. breve* k1 gene into the *E. coli* genome

Integrating the desired DNA into the cell's genome may enhance the stability of gene expression over time and under a variety of fermentation conditions. However, the location of integration may affect gene expression level and, ultimately, the effectiveness of the desired enzyme activity.

Integration of the k1 expression cassette (NPR promoter-signal peptide-k1 gene) into the genome of *E. coli* (strain FM5) and the demonstration of utility by the use of an α(1,6)-linked glucose substrate was accomplished by first cloning into the plasmid pKD3 (Datsenko and Wanner, *Proc. Natl. Acad. Sci.* 97:6640-6645 (2000)). The host aldA (aldehyde dehydrogenase A) and aldB (aldehyde dehydrogenase B) genomic sites were chosen for integration. PCR primers were designed

that had homology to the plasmid pKD3, aldA or aldB and k1 gene sequences (SEQ ID NOs:20 through 23).

PCR amplification was performed by the protocol described in Example 3. PCR products resulting from a reaction with the primers SEQ ID NOs. 21-23 and the plasmid pKD3 containing the k1 expression cassette were isolated, ligated and transformed into *E. coli* (FM5). Cells containing the integrated k1 expression cassette were selected by growth on LB medium containing chloramphenicol. Chloramphenicol positive colonies were tested for the presence of the k1 gene by PCR reaction, using the primers SEQ ID NO:7 and SEQ ID NO:8.

FM5 lines containing the integrated k1 expression cassette were further tested for activity by growth analysis in media containing isomaltose as the sole carbohydrate source. Chloramphenicol and PCR positive colonies were inoculated into TM2 medium (see Example 7) with 0.5 % isomaltose (w/v) and grown at 35 °C. Samples were removed at various time points and characterized for cell mass accumulation by optical density (A600nm) and isomaltose consumption (by HPLC, see General Methods).

Table 12 below demonstrates that FM5 cells alone do not metabolize isomaltose when provided as the sole carbohydrate source. This is shown by the low level of cell mass accumulation when grown in TM2 medium with 0.5 % isomaltose. Low-level growth of the negative line FM5 was observed, but due only to a small amount of the fermentable sugar maltose contaminating the isomaltose source material (Sigma, St. Louis, MO). Cells containing the integrated K1 expression cassette grew at a much higher rate and to a higher final optical density following the 25 hr time period. A PCR-positive colony containing the k1 expression cassette integrated at the aldA site was designated A2-3. Colonies, positive by PCR, containing the k1 expression cassette integrated at the aldB site were designated B1-1 and B1-2.

Table 12

Cell mass accumulation (A600nm)

Time (hours)	FM5	FM5-A2-3	FM5-B1-1	FM5-B1-2
0	0.02	0.02	0.02	0.02
3	0.66	0.72	0.76	0.75
6	2.75	3.17	6.60	6.01
8	3.34	4.50	10.40	9.92
11	3.72	8.34	10.41	10.10
25	3.66	10.16	11.10	10.78

Isomaltose consumption by cells containing the integrated K1 expression cassette was also compared to the FM5 negative control line by HPLC analysis. The data in Table 13 demonstrate that the K1 expression cassette is active following integration and allows cells to completely utilize available sugar containing α(1,6)-linked glucose, compared to the negative control which does not utilize this carbohydrate. The data also show that isomaltose is not consumed at the same rate in lines where the gene has been integrated into the aldA, as compared to the aldB, sites.

Table 13

Isomaltose Consumption (g/L)

Time (hours)	FM5	FM5-A2-3	FM5-B1-1	FM5-B1-2
0	5.56	5.46	5.36	5.31
3	5.52	5.35	5.31	5.30
6	5.60	4.73	1.81	1.78
8	5.48	3.64	0	0
11	5.77	1.34	0	0
25	5.55	0	0	0